

# Breve Introducción a R

*Juan Carlos Castillo*

## Notas introductorias

- R Corresponde más a un marco de análisis estadístico que a un programa estadístico, con una fuerte orientación a ciencia de datos (Data Science)
- Gratuito y de código abierto
- Actualmente se combina con una serie de herramientas de ciencia de datos para facilitar reportes y reproducibilidad de los análisis
- El registro de los análisis queda en formato de texto plano, por lo tanto es independiente de una plataforma para poder editarlo, y además permite un control eficiente de versiones (por ejemplo vía Git).
- Los análisis operan en base a paquetes o librerías
- Actualmente existen más de 3000 librerías disponibles
- Particularidad: el análisis se orienta a objetos (detalle más adelante)

## 1. Instalación R / Rstudio

- Visitar la página de CRAN (Comprehensive R Archive Network), <http://cran.r-project.org/>
- Seleccionar versión según sistema operativo (Ej. Linux)
- Instalar “base”
- En el caso de Windows, el programa R GUI (Graphical User Interface) se agrega a la lista de programas y aparece como de acceso desde el escritorio. Es la interfaz de R por defecto
- Nota: Existen otras interfaces gráficas (GUIs) para trabajar con R, como Commander o Java GUI for R (Jaguar), Deducer, R-Studio, etc.
- Interfaz recomendada: R Studio, <https://www.rstudio.com/>
- Actualización:
  - Bajar e instalar nueva versión
  - Copiar librerías de carpeta antigua a la nueva
  - Actualizar librerías (update packages)

## 2. Trabajando en R

### Bases

- RStudio posee 4 ventanas (panes): las dos principales son el editor (source) y consola; las otras dos tienen relación con entorno, paquetes, plots, etc.
- El orden como se presentan estas cuatro “panes” se puede cambiar en Tools/Options/Pane Layout
- Se pueden ingresar comandos directamente en la consola, por ejemplo:

```
4+1  
4>1
```

- Los comandos se escriben en el prompt (>), y se ejecutan con la tecla **enter**

- El número entre paréntesis cuadrado (ej. [1]) indica el orden de aparición de los resultados
- Sin embargo, la manera habitual de trabajar es de un archivo de código, donde quedan registrados los comandos (análogo a do file de Stata)
- También existen algunas opciones de menú, las que tienen relación con edición y la configuración del programa (no con el análisis de datos)
- Para salir de R , cerrar ventana o ejecutar `q()`

## Trabajando desde un archivo de código (script)

- File - new script
- Archivo en que se ingresan los comandos correspondientes a un análisis específico, los cuales pueden ser guardados y ejecutados posteriormente
- Para correr los comandos desde el editor, posicionar el cursor en la línea respectiva y luego “ctrl r” o “F5”, o con el icono de ejecución
- Para grabar scripts: File - save as
- Por defecto graba con extensión `.R`, pero es sólo un archivo de formato simple (txt) que se puede abrir con cualquier editor de texto (ej. Block de notas).
- Para abrir script grabado: File - open script
- Caracteres especiales

```
# Comentarios (AltGr 3) , no se ejecutan
+ Sigue el comando en la próxima línea
; Para escribir más de una función en la misma línea de comandos
```

## Librerías

- Conjunto de funciones que tienen una relación entre ellas y que usualmente vienen acompañadas de ficheros de ayuda (documentación)
- Algunas librerías vienen preinstaladas, otras específicas hay que instalarlas de acuerdo a las necesidades del usuario
- Para conocer la lista de librerías instaladas: `library()`
- Para instalar: `install.packages(librería)`, en el caso que se sepa el nombre específico de la librería que se quiera instalar. O mediante menú : Packages - Install package(s)
- Las librerías se instalan sólo 1 vez y quedan guardadas en una carpeta local, pero deben ser cargadas si se quieren utilizar en la sesión de trabajo con `library("library")`
- Para explorar librerías disponibles: <http://cran.r-project.org/>, organizadas por área en Task Views
- Ej: instalar librería psy

```
install.packages("psy")
library(psy)
? psy # Ayuda de la libreria
```

## Instalación/carga de librerías con pacman

- Para evitar el tener que instalar/cargar librerías, la mejor solución es utilizar la librería **pacman**
- Se instala, y luego en las siguientes sesiones se utiliza el siguiente código con las librerías a utilizar (por ejemplo, lavaan)

```
# install.packages("pacman") # solo la primera vez
pacman::p_load(lme4)
```

- Si la librería está instalada, solo la carga; si no, la instala y la carga.

## Objetos

- R es un programa orientado a objetos, los que son creados por funciones, que en su forma más general sería: `Objeto <- función` o de manera equivalente `Objeto = función`
- Diferentes tipos de objetos: vectores, factores, matrices, marco/base de datos (entre otros)
- Objetos simples:

```
x = 5 # el número 5 es asignado al objeto x
x
```

```
## [1] 5
```

```
## [1] 5
a = "hoy"
a
```

```
## [1] "hoy"
```

```
## [1] "hoy"
```

## Vectores

- Objeto unidimensional constituido por elementos del mismo tipo

```
edad = c(50,1,25,6) # c es por "concatenate"
edad
```

```
## [1] 50  1 25  6
```

```
alumnos = c("juan", "simón", "maría", "sonia")
alumnos
```

```
## [1] "juan" "simón" "maría" "sonia"
```

```
hasta20 = 1:20 #genera una secuencia de números del 1 al 20
hasta20
```

```
## [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
```

```
a = seq(from = -5, to = 5, by = 1.5)
a
```

```
## [1] -5.0 -3.5 -2.0 -0.5  1.0  2.5  4.0
```

```
tresdos = rep(2,3)
tresdos
```

```
## [1] 2 2 2
```

Ejemplo de operaciones con vectores numéricos

```
mean(edad)
```

```
## [1] 20.5
```

```
summary(edad)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00   4.75   15.50   20.50   31.25   50.00
```

*Algunas otras funciones*

Función	Resultado
median ( )	Mediana del vector x
mean ( )	Media del vector
quantile ( )	Cuantiles
IQR ( )	Rango intercuartil
range (x)	Rango del vector x
sd (x)	Desviación estándar
var ( )	Varianza de los elementos
summary ( )	Resumen descriptivos

## Factores

- Modo que utiliza R para almacenar variables categóricas

```
sexo <- c(rep("mujer", 700), rep("varon", 569)) # crea vector de caracteres
head(sexo) # muestra 6 primeros valores
```

```
## [1] "mujer" "mujer" "mujer" "mujer" "mujer" "mujer"
```

```
#Para convertir el vector en factor:
sexo <- as.factor(sexo)
levels(sexo)
```

```
## [1] "mujer" "varon"
```

```
table(sexo)
```

```
## sexo
## mujer varon
##    700    569
```

## Matrices

- objeto bidimensional constituido por filas y columnas de elementos del mismo tipo

```
x <- matrix(1:9,3,3)
x
```

```
##      [,1] [,2] [,3]
## [1,]    1    4    7
## [2,]    2    5    8
## [3,]    3    6    9
```

```
x <- matrix(1:8,2,4,byrow = F) # genera una matriz con 2 filas y 4 columnas que se irá completando por
x
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    3    5    7
## [2,]    2    4    6    8
```

```
x <- matrix(1:8,2,4,byrow = T) # genera la matriz completándola por filas
x
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    2    3    4
## [2,]    5    6    7    8
```

## Marco de datos (dataframe)

- Estructura más común en R para el análisis de datos
- Consiste en una matriz donde las columnas pueden tener datos almacenados en distintos modos , numéricos y categóricos
- Se pueden concebir como conjuntos de datos donde las líneas representan casos y las columnas variables
- Las variables pueden ser de distinto tipo, pero todos los datos referidos a una misma variable son del mismo modo
- Creación de dataframes a partir de matrices

```
x <- matrix(1:9,3,3)
x
```

```
##      [,1] [,2] [,3]
## [1,]    1    4    7
## [2,]    2    5    8
## [3,]    3    6    9
```

```
x <- as.data.frame(x)
x
```

```
##   V1 V2 V3
## 1  1  4  7
## 2  2  5  8
## 3  3  6  9
```

```
# Modificar nombres de filas y columnas
names(x) <- c("Sociología", "Psicología", "Derecho")
row.names(x) <- c("Juan", "Maria", "Pedro")
x
```

```
##      Sociología  Psicología  Derecho
## Juan          1          4          7
## Maria         2          5          8
## Pedro         3          6          9
```

```
# Operaciones simples
summary(x)
```

```
##      Sociología  Psicología  Derecho
## Min.   :1.0    Min.   :4.0    Min.   :7.0
## 1st Qu.:1.5    1st Qu.:4.5    1st Qu.:7.5
## Median :2.0    Median :5.0    Median :8.0
## Mean   :2.0    Mean   :5.0    Mean   :8.0
## 3rd Qu.:2.5    3rd Qu.:5.5    3rd Qu.:8.5
## Max.   :3.0    Max.   :6.0    Max.   :9.0
```

```
View(x)
```

```
mean(x$Derecho)
```

```
## [1] 8
```

```
# Guardar datos
write.table(x,file="data1.csv",sep="," ,row.names=FALSE,col.names=TRUE)
```

- Creación de dataframes a partir de vectores

```
edad <- c(50,1,25,6) # c es por "concatenate"
alumnos <- c("juan", "simón", "maría", "sonia")
curso=data.frame(Edad=edad, Alumnos=alumnos)
curso
```

```
##      Edad Alumnos
## 1     50     juan
## 2      1     simón
## 3     25     maría
## 4      6     sonia
```

```
View(curso) # visualizar datos
```

```
mean(curso$Edad) # promedio de la variable Edad del dataframe curso
```

```
## [1] 20.5
```

```
attach(curso)      # establece que curso es el dataframe por defecto para funciones
mean(edad)         # curso attached, no es necesario indicar "curso$"
```

```
## [1] 20.5
```

```
# Agregar variable
curso$genero=c(0,0,1,1) # 1=mujer

# Descriptivos generales
names(curso)
```

```
## [1] "Edad"      "Alumnos" "genero"
```

```
summary(curso)
```

```
##      Edad      Alumnos      genero
## Min.   : 1.00   juan :1   Min.    :0.0
## 1st Qu.: 4.75   maría:1  1st Qu.:0.0
## Median :15.50   simón:1  Median :0.5
## Mean   :20.50   sonia:1  Mean   :0.5
## 3rd Qu.:31.25           3rd Qu.:1.0
## Max.   :50.00           Max.    :1.0
```

## Lectura de base de datos

Indicándole la ruta donde se encuentra la base de datos

```
data = read.csv("data1.csv", header=T)
```

Método alternativo (sin indicarle la ruta donde se encuentra la base de datos)

```
data = read.csv(file.choose(), header=T, sep=", ")
```

- Recomendación: establecer directorio de trabajo al comienzo del script (donde se buscan y guardan los archivos)\*

```
getwd() # obtener directorio de trabajo actual
setwd() # establecer directorio de trabajo
# Ej: windows
setwd("C:/Documents and Settings/jcastillo/Misdocumentos/proyecto1")
# Ej: linux
setwd("/media/ntfs/Dropbox/cursos/isuc/multinivel")
```

Nota: las carpetas de ruta en Windows van separadas por slash (/), en Linux o Mac con backslash (\).

## Exploración de base de datos

```
attach(data) # facilita la operación con la base de datos data  
names(data) # muestra los nombres de las variables de la base de datos
```

```
## [1] "Sociología" "Psicología" "Derecho"
```

```
View(data) #muestra en detalle toda la base de datos
```